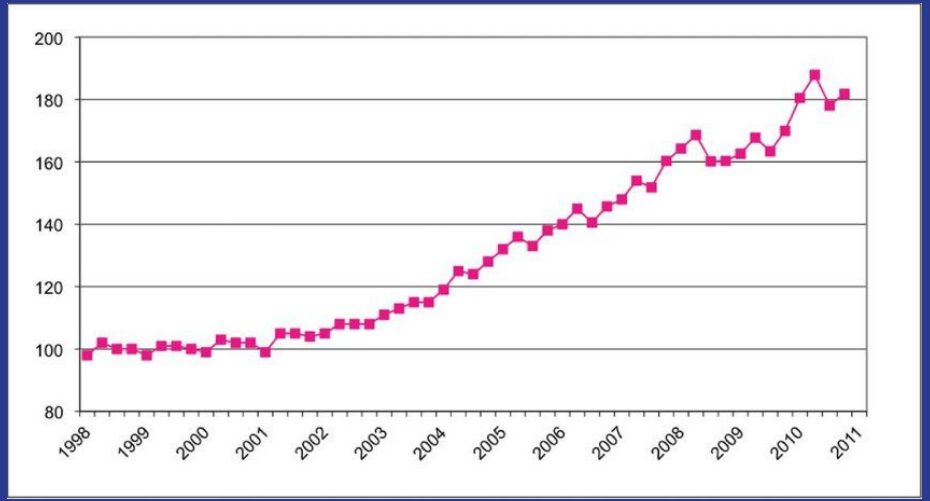
Beijing House Price (2016)

Group 1: Cinny Lin (ycl461), Yihan Xu (yx1708), Yizhou Lu (yl5438)

Why Housing Price?

House prices

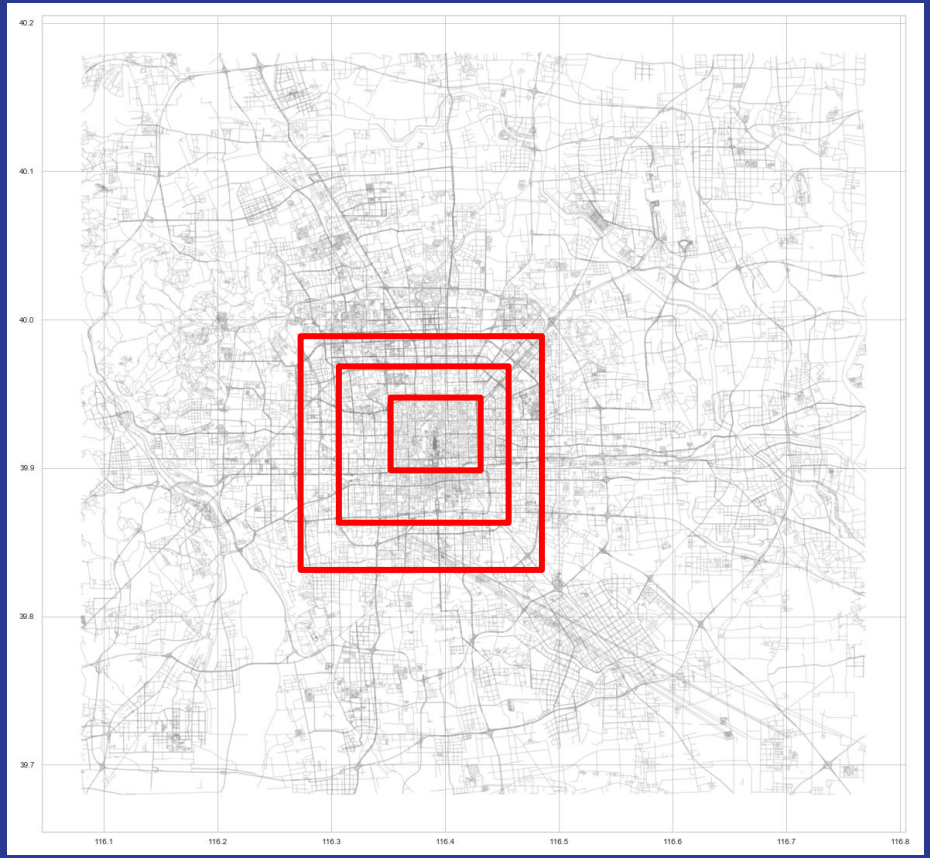
rising in China



Why Beijing?

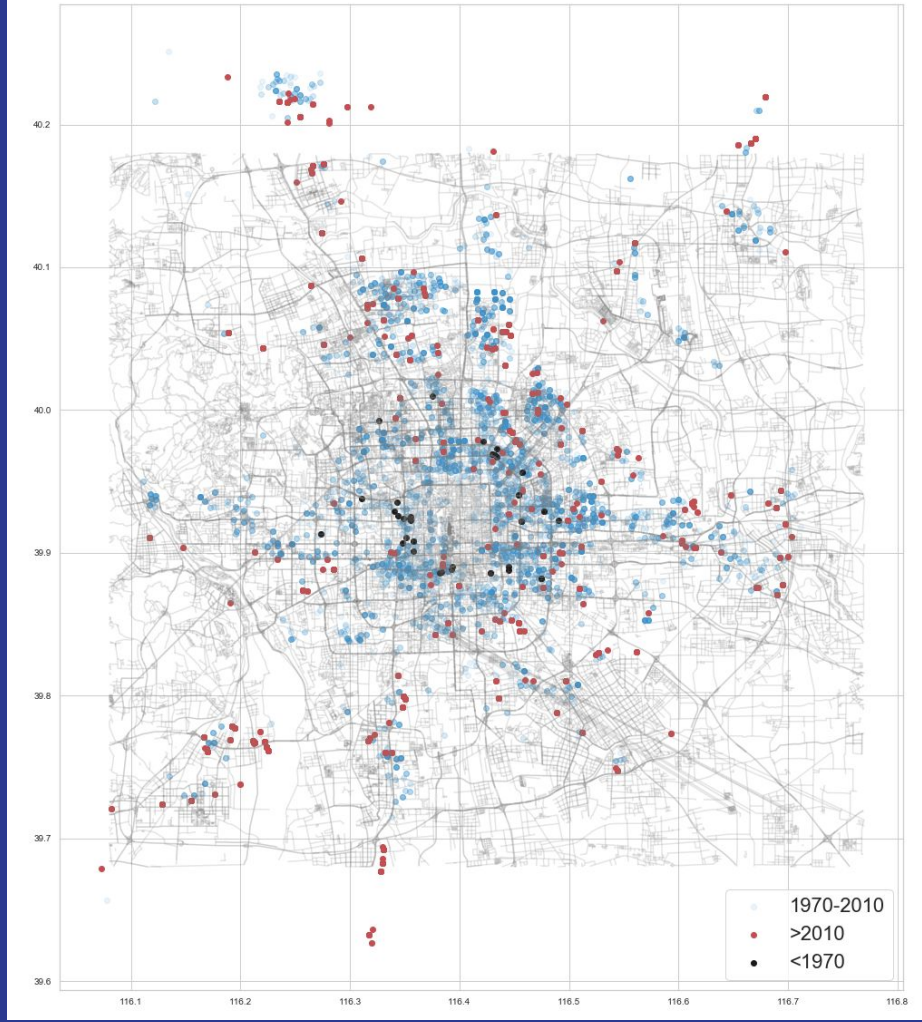
Location

concentric city layout



Age

coexistence of
modern and ancient houses



Overview

Dataset

10,682 observations of housing price in Beijing in 2016, from Lianjia.com

Question

What are the key factors that influence per unit housing prices in Beijing?

Hypothesis

Location and **age** are determinant factors.

Variable Description and Relationships

Basic Descriptive Statistics

N = 10331, Beijing, China, 2016

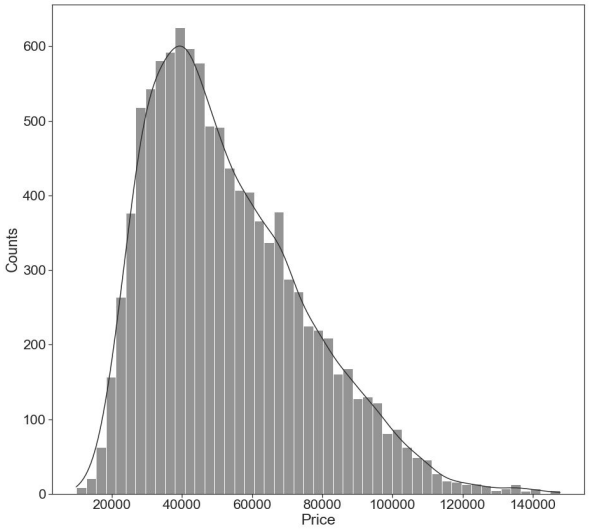
	Min	LQ	Med	UQ	Max	Mean	Std.dev	Skewness
price	9841	35891	48659	67151	147797	53082.72	22340.22	0.81
square	13.7	59	77.73	105.91	458	89.79	46.3	2.17
Lng	116.07	116.34	116.41	116.47	116.71	116.41	0.1	0.07
Lat	39.63	39.9	39.94	40	40.25	39.95	0.095	0.29
livingRoom	0	2	2	3	9	2.1	0.86	0.86
drawingRoom	0	1	1	2	5	1.19	0.57	0.36
kitchen	0	1	1	1	2	0.99	0.12	-3.54
bathRoom	0	1	1	1	6	1.24	0.53	2.35
constructionTime	1944	1994	2002	2006	2016	1999.89	8.97	-0.92
ladderRatio	0.01	0.25	0.33	0.5	2	0.38	0.19	2.07
elevator	0	0	1	1	1	0.61	0.49	-0.43
fiveYearsProperty	0	0	1	1	1	0.62	0.49	-0.49
subway	0	0	1	1	1	0.62	0.48	-0.51
*buildingType*1	0	0	0	1	1	0.28	0.45	0.99
*buildingType*2	0	0	0	0	1	0.0004	0.02	45.42
*buildingType*3	0	0	0	0	1	0.19	0.39	1.58
*buildingType*4	0	0	1	1	1	0.52	0.50	-0.09
*renovationCondition*1	0	0	0	0	1	0.07	0.25	71.85
*renovationCondition*2	0	0	0	0	1	0.03	0.17	5.44
*renovationCondition*3	0	0	0	1	1	0.37	0.48	0.54
*renovationCondition*4	0	0	1	1	1	0.53	0.5	-0.12
*buildingStructure*1	0	0	1	1	1	0.53	0.5	71.85
*buildingStructure*2	0	0	0	1	1	0.33	0.47	0.71
*buildingStructure*3	0	0	0	0	1	0.0004	0.02	45.42
*buildingStructure*4	0	0	0	0	1	0.04	0.21	4.37
*buildingStructure*5	0	0	0	0	1	0.001	0.03	20.29
*buildingStructure*6	0	0	1	1	1	0.62	0.49	-0.5

Basic Descriptive Analysis

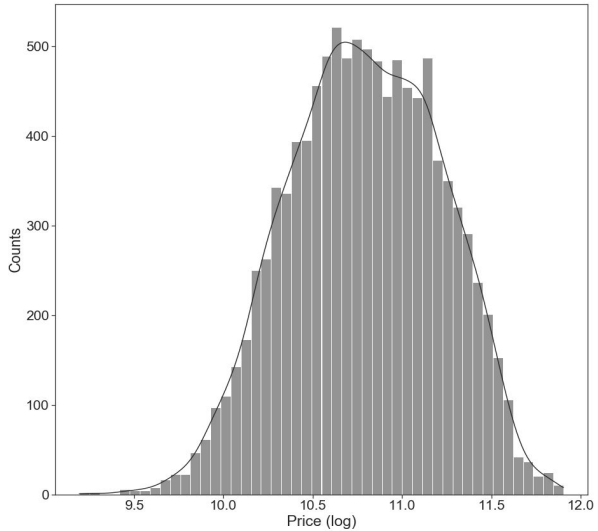
price

house price per square unit

price



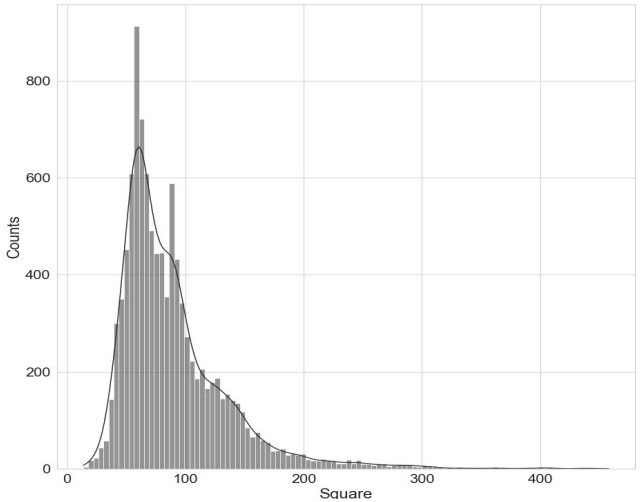
log(price)



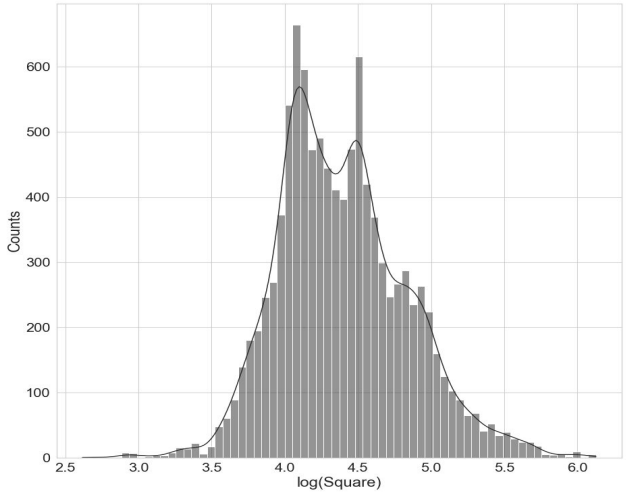
square

total house size

square



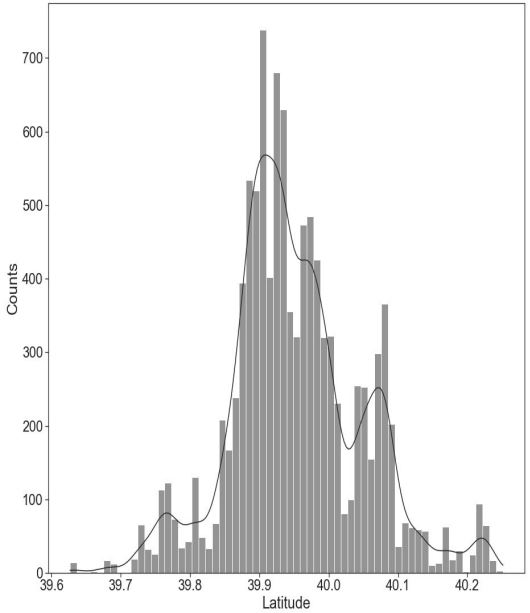
log(square)



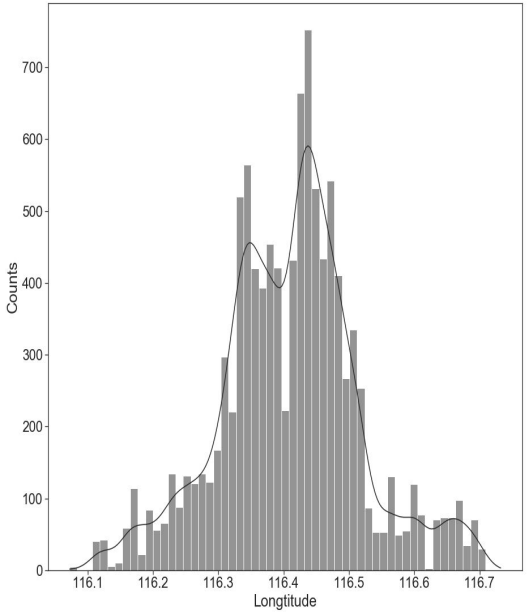
location

latitude and longitude

latitude



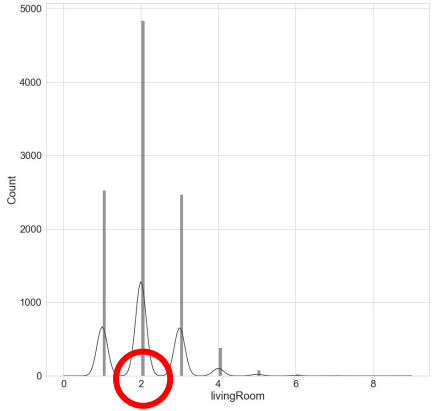
longitude



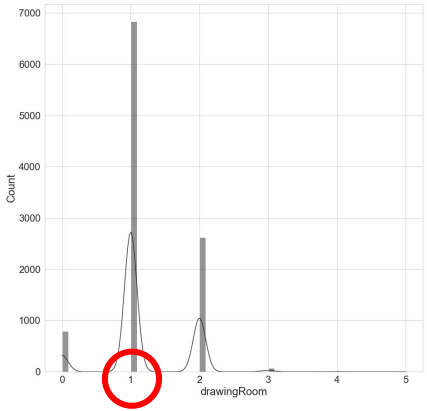
rooms

number of rooms in a house

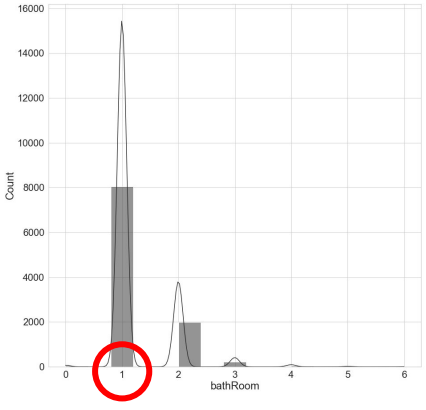
living room



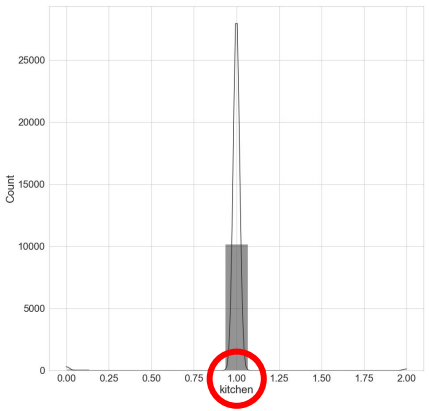
drawing room



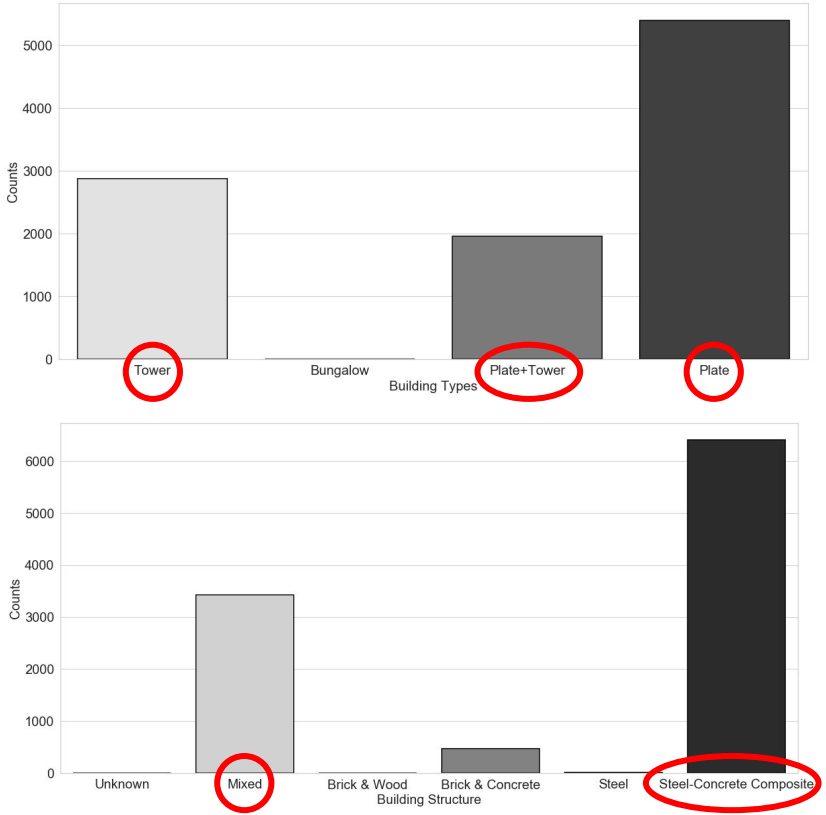
bathroom



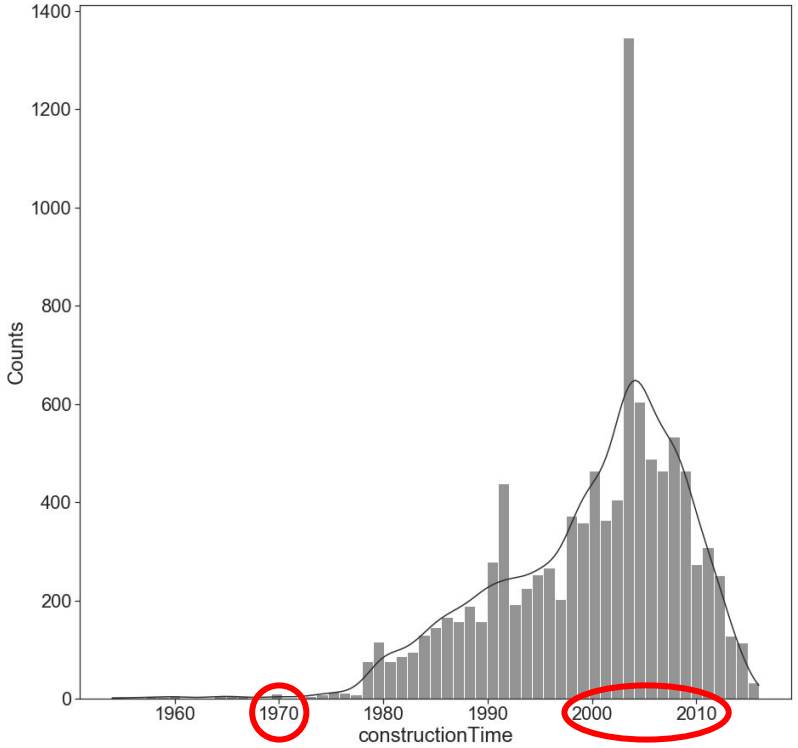
kitchen



building type & structure

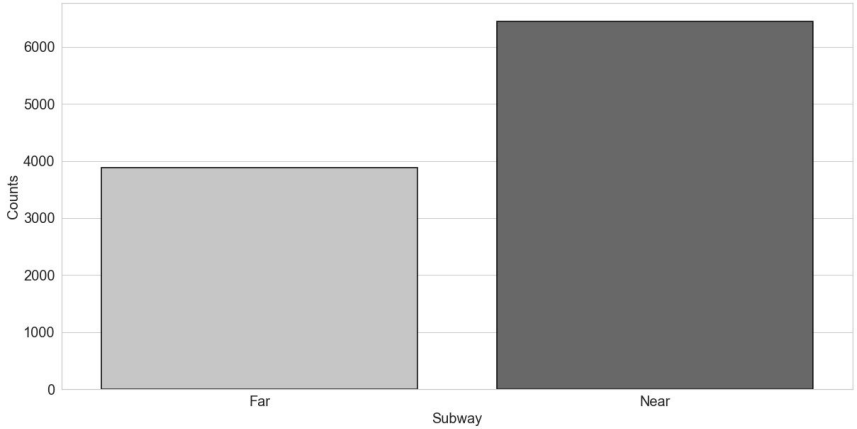


construction time

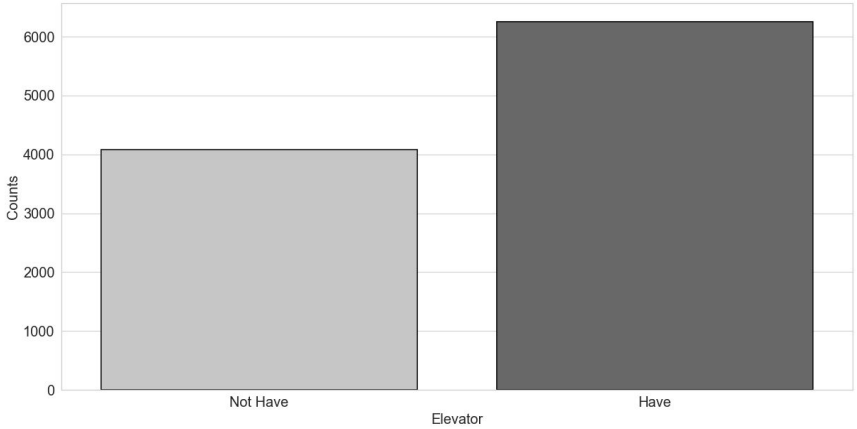


transportation

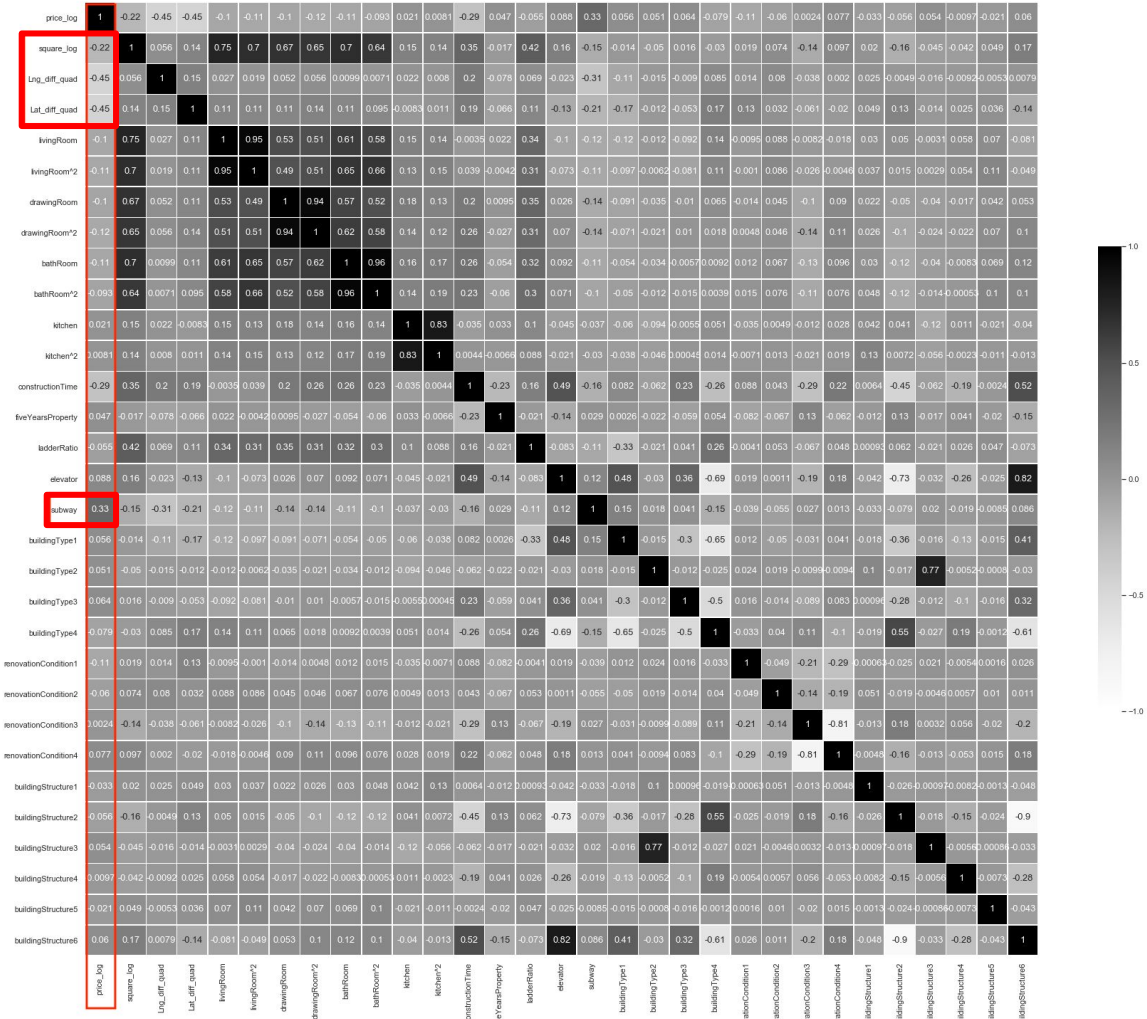
subway

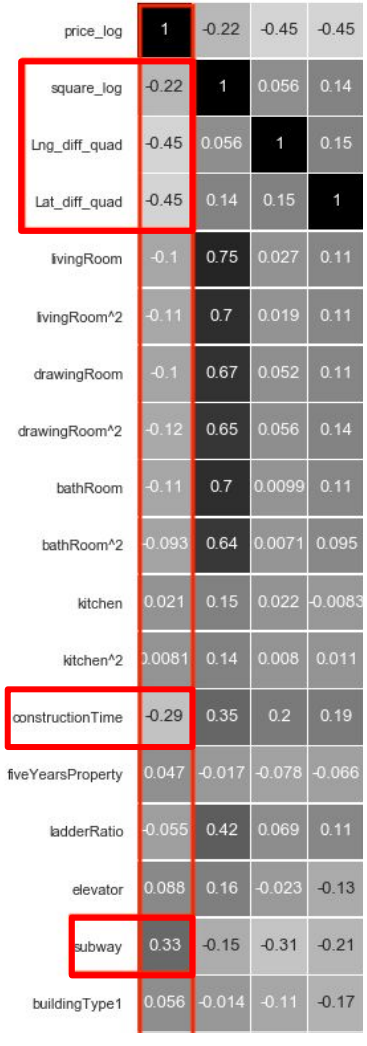


elevator



Correlation Matrix

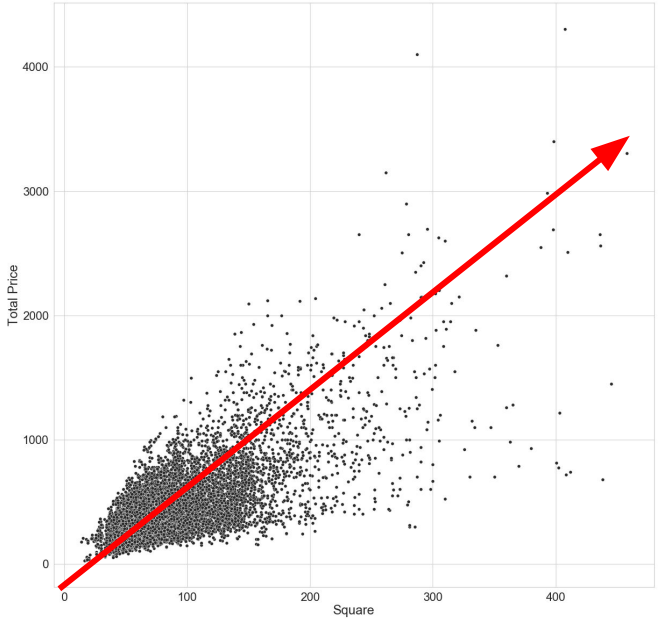

price_log
square_log
Lng_diff_quad
Lat_diff_quad
livingRoom
livingRoom^2
drawingRoom
drawingRoom^2
bathRoom
bathRoom^2
kitchen
kitchen^2
constructionTime
fiveYearsProperty
ladderRatio
elevator
subway
buildingType1
buildingType2
buildingType3
buildingType4
renovationCondition1
renovationCondition2
renovationCondition3
renovationCondition4
buildingStructure1
buildingStructure2
buildingStructure3
buildingStructure4
buildingStructure5
buildingStructure6
1.0
0.5
0.0
−0.5
−1.0


price_log 1 -0.22 -0.45 -0.45
square_log -0.22 1 0.056 0.14
Lng_diff_quad -0.45 0.056 1 0.15
Lat_diff_quad -0.45 0.14 0.15 1
livingRoom -0.1 0.75 0.027 0.11
livingRoom^2 -0.11 0.7 0.019 0.11
drawingRoom -0.1 0.67 0.052 0.11
drawingRoom^2 -0.12 0.65 0.056 0.14
bathRoom -0.11 0.7 0.0099 0.11
bathRoom^2 -0.093 0.64 0.0071 0.095
kitchen 0.021 0.15 0.022 -0.0083
kitchen^2 0.0081 0.14 0.008 0.011
constructionTime -0.29 0.35 0.2 0.19
fiveYearsProperty 0.047 -0.017 -0.078 -0.066
ladderRatio -0.055 0.42 0.069 0.11
elevator 0.088 0.16 -0.023 -0.13
subway 0.33 -0.15 -0.31 -0.21
buildingType1 0.056 -0.014 -0.11 -0.17

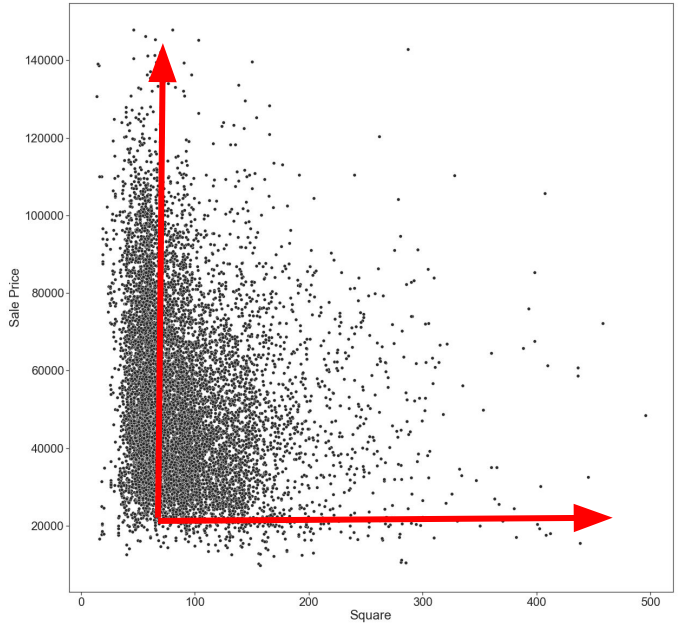
Correlation Analysis

price or total price?

total price



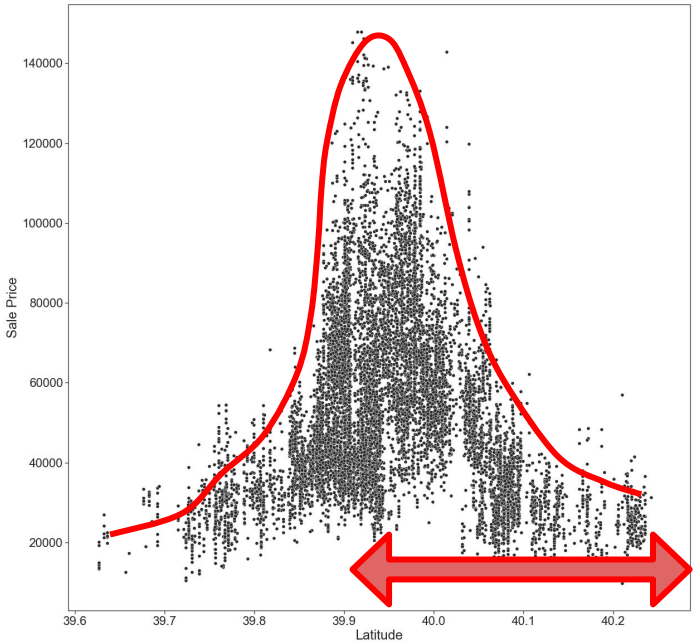
price



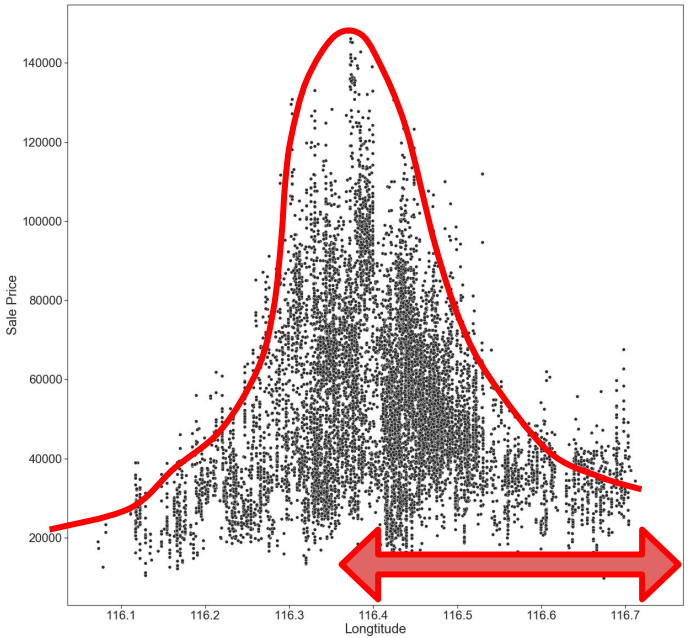
location

latitude and longitude

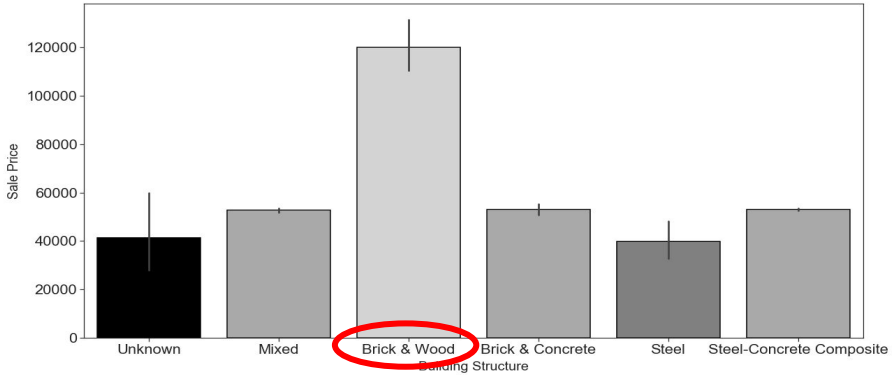
latitude



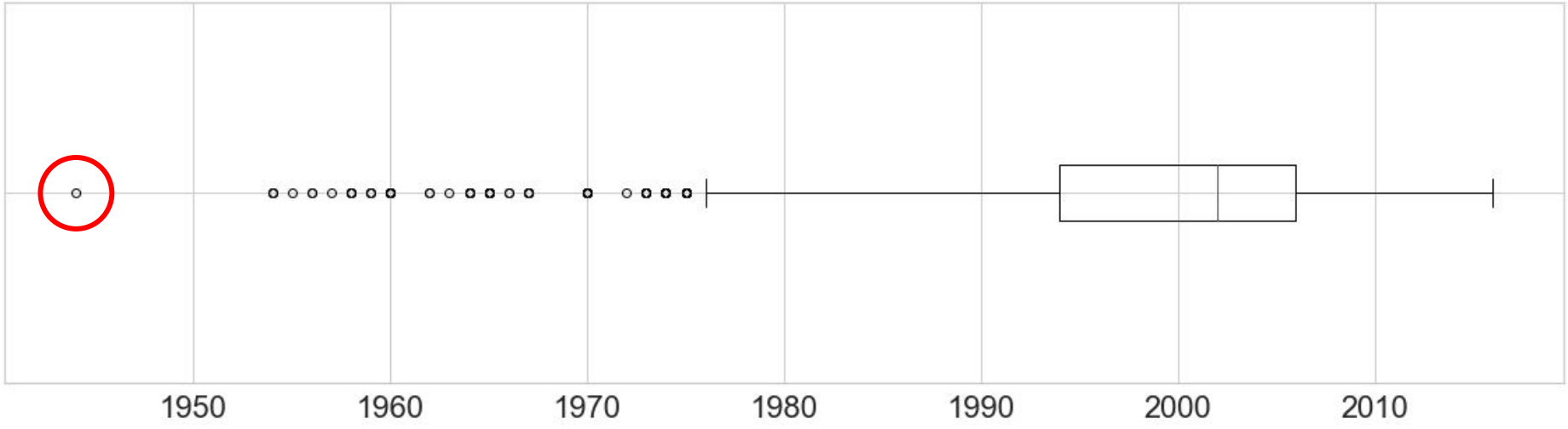
longitude



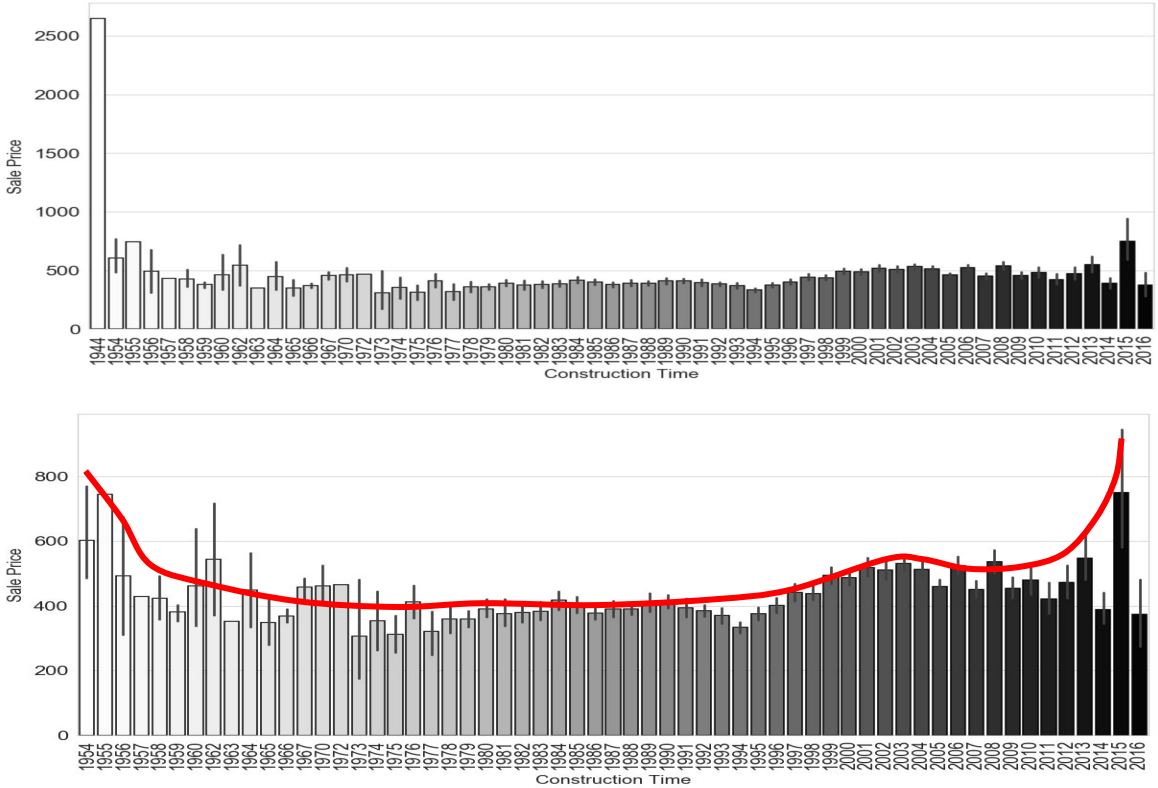
building type & structure



construction time

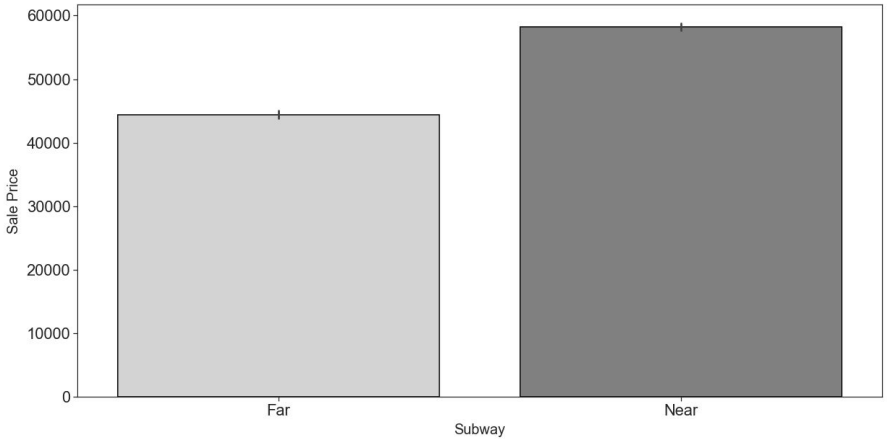


construction time

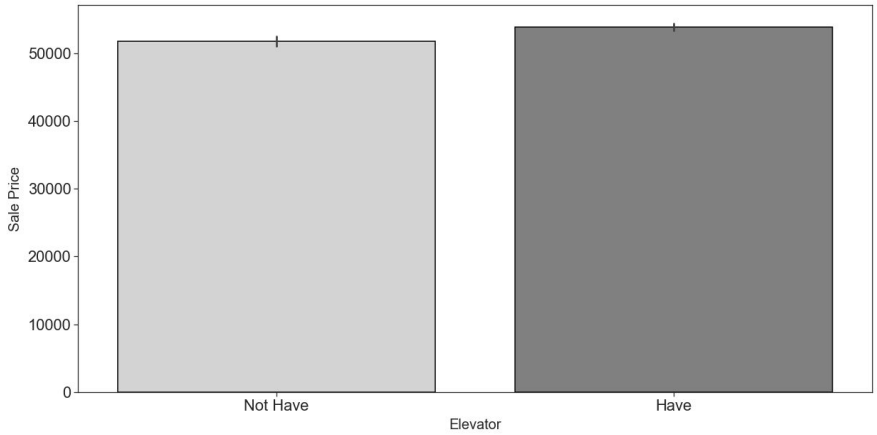


transportation

subway



elevator



Inferential Analysis

Model Construction

Functional Form

- Based on variable description
  - **log(price)** → dependent variable
  - **log(square)** → independent variable
- Based on literature review
  - $(A * age + B * age^2)$
  - $(A * n + B * n^2)$ → number of rooms (n)

First Iteration

First Iteration Regression

Source	SS	df	MS	Number of obs	=	10,330
				F(27, 10302)	=	306.05
Model	824.017038	27	30.5191496	Prob > F	=	0.0000
Residual	1027.31887	10,302	.099720333	R-squared	=	0.4451
				Adj R-squared	=	0.4436
Total	1851.33591	10,329	.179236703	Root MSE	=	.31579

price_log	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Lng_diff_quad	-7.801252	.1845434	-42.27	0.000	-8.162993	-7.439511
Lat_diff_quad	-5.981557	.1581843	-37.81	0.000	-6.291629	-5.671485
square_log	-.3232739	.0155626	-20.77	0.000	-.3537795	-.2927683
livingRoom	.1331211	.0148177	8.98	0.000	.1040755	.1621667
livingRoom2	-.0173575	.002662	-6.52	0.000	-.0225755	-.0121394
drawingRoom	.1285189	.0174101	7.38	0.000	.0943917	.162646
drawingRoom2	-.0236394	.0061725	-3.83	0.000	-.0357388	-.01154
kitchen	.098927	.0550441	1.80	0.072	-.00897	.2068241
kitchen2	.0060342	.0300601	0.20	0.841	-.0528895	.0649579
bathRoom	-.0552431	.0238766	-2.31	0.021	-.1020459	-.0084403
bathRoom2	.030211	.0060237	5.02	0.000	.0184033	.0420187
constructionTime_diff	.0109332	.0013903	7.86	0.000	.0082079	.0136585
constructionTime_diff2	-.0000485	.0000298	-1.63	0.103	-.0001069	9.84e-06
ladderRatio	.1684176	.0193751	8.69	0.000	.1304388	.2063965
elevator	.1502676	.0129221	11.63	0.000	.1249377	.1755975
fiveYearsProperty	-.0193993	.0067672	-2.87	0.004	-.0326643	-.0061344
subway	.0927824	.0070136	13.23	0.000	.0790345	.1065304
buildingType2	.6230491	.2747338	2.27	0.023	.0845173	1.161581
buildingType3	.0730704	.0096089	7.60	0.000	.0542351	.0919056
buildingType4	.0971986	.010536	9.23	0.000	.076546	.1178511
renovationCondition2	.0377757	.0215294	1.75	0.079	-.004426	.0799774
renovationCondition3	.0234049	.0131818	1.78	0.076	-.002434	.0492438
renovationCondition4	.104152	.0126615	8.23	0.000	.079333	.128971
buildingStructure2	.0493564	.2239212	0.22	0.826	-.3895726	.4882855
buildingStructure3	0	(omitted)				
buildingStructure4	.0356348	.2243508	0.16	0.874	-.4041363	.4754059
buildingStructure5	.03659	.2423199	0.15	0.880	-.4384041	.511584
buildingStructure6	.1070451	.2240259	0.48	0.633	-.3320892	.5461795
_cons	11.41771	.2329978	49.00	0.000	10.96099	11.87443

Second Iteration

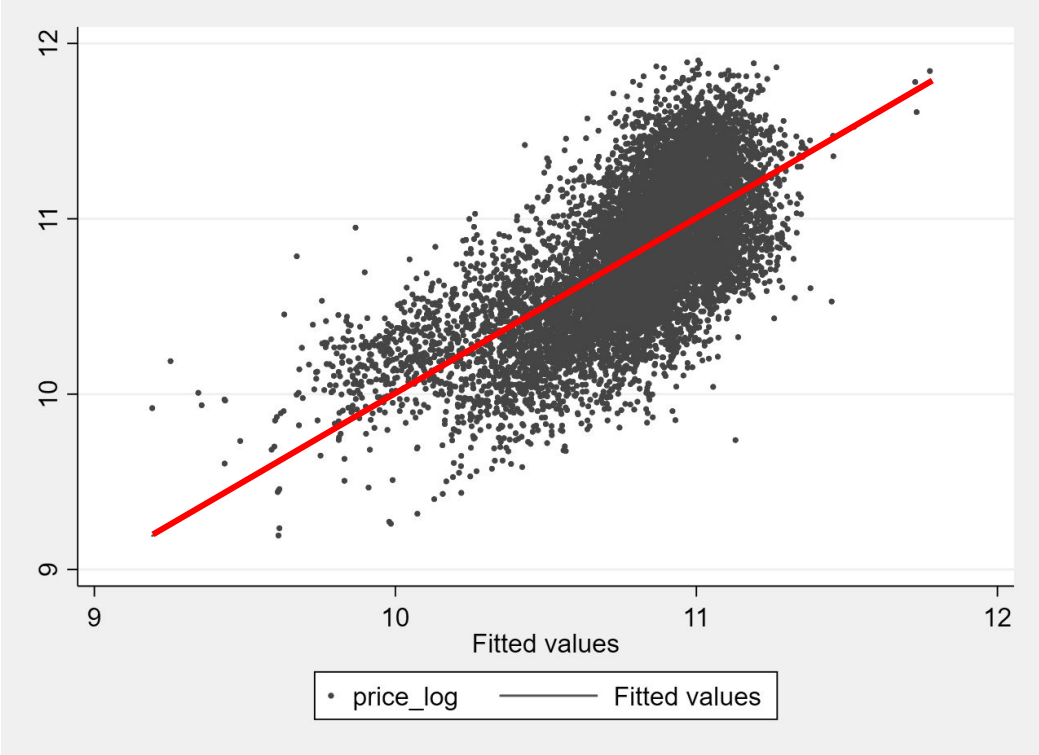
Regression Model Refined

Source	SS	df	MS		
Model	819.364779	18	45.5202655	Number of obs	= 10,330
Residual	1031.97113	10,311	.100084485	F(18, 10311)	= 454.82
				Prob > F	= 0.0000
Total	1851.33591	10,329	.179236703	R-squared	= 0.4426
				Adj R-squared	= 0.4416
				Root MSE	= .31636

price_log	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Lng_diff_quad	-7.768248	.1839029	-42.24	0.000	-8.128733	-7.407763
Lat_diff_quad	-6.131044	.1561888	-39.25	0.000	-6.437204	-5.824883
square_log	-.3235396	.015573	-20.78	0.000	-.3540658	-.2930135
livingRoom	.1338677	.0147842	9.05	0.000	.1048878	.1628476
livingRoom2	-.0174411	.0026519	-6.58	0.000	-.0226392	-.0122429
drawingRoom	.138355	.0170525	8.11	0.000	.1049288	.1717812
drawingRoom2	-.0258867	.0060764	-4.26	0.000	-.0377976	-.0139757
bathRoom	-.0487354	.0236405	-2.06	0.039	-.0950754	-.0023953
bathRoom2	.0293854	.0059681	4.92	0.000	.0176868	.0410839
constructionTime_diff	.0084452	.0005022	16.82	0.000	.0074608	.0094295
ladderRatio	.1663154	.0193573	8.59	0.000	.1283714	.2042595
elevator	.1850089	.0102149	18.11	0.000	.1649856	.2050322
fiveYearsProperty	-.0172248	.006644	-2.59	0.010	-.0302482	-.0042013
subway	.0938314	.0070199	13.37	0.000	.080071	.1075918
buildingType2	.5083676	.1592971	3.19	0.001	.1961144	.8206208
buildingType3	.071238	.009554	7.46	0.000	.0525103	.0899657
buildingType4	.0892781	.0102589	8.70	0.000	.0691687	.1093876
renovationCondition4	.0843148	.0064586	13.05	0.000	.0716546	.0969749
_cons	11.62013	.0580308	200.24	0.000	11.50638	11.73388

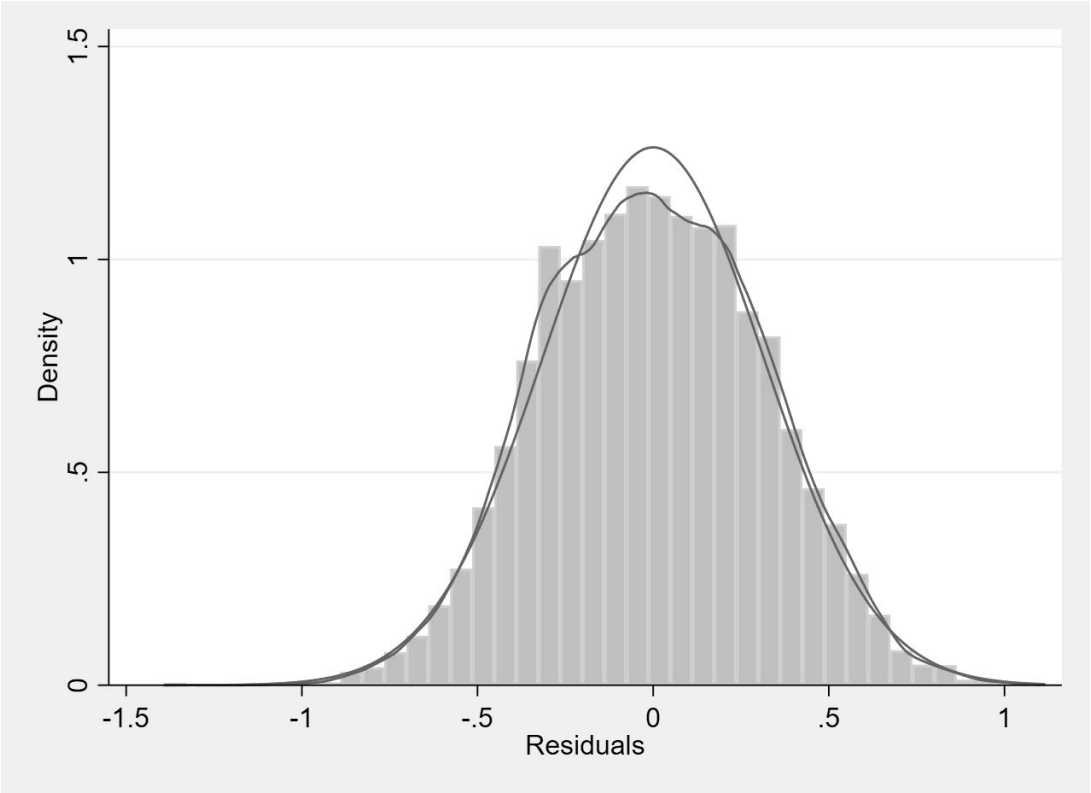
Model Diagnostic Test

Fitted Values V.S. Observed Values



1) *Linear in parameters*
2) *Random sampling*
3) *No perfect collinearity*
4) ***Zero-conditional mean***
5) *Homoscedasticity*
6) *Normality*

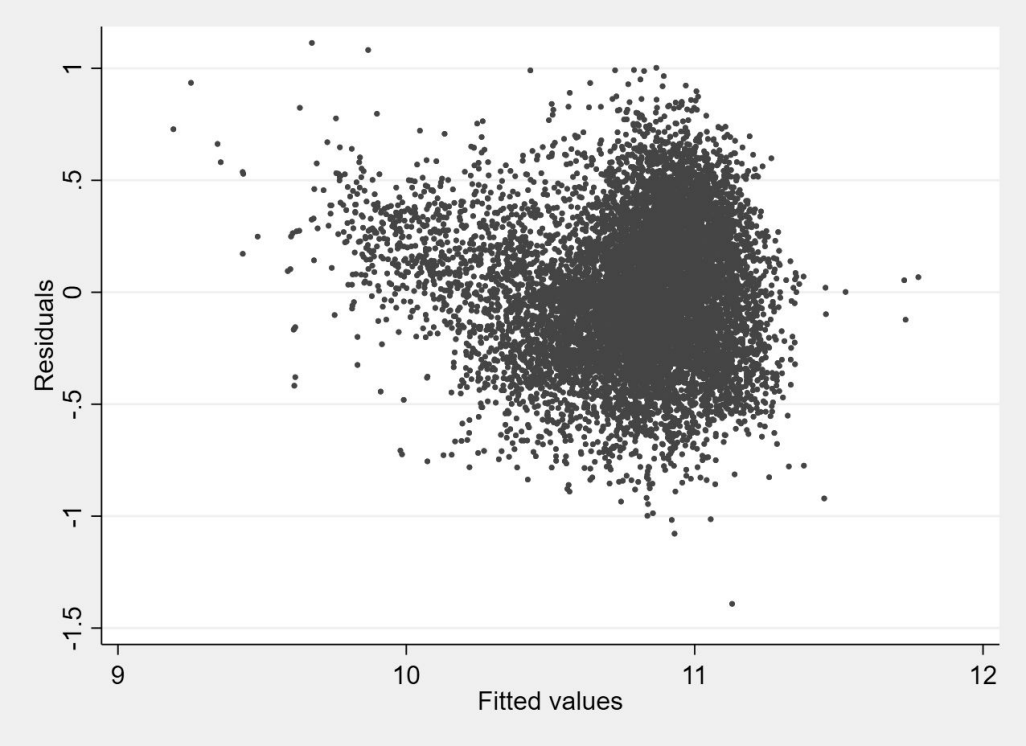
Normal Distribution of Residuals



1) *Linear in parameters*
2) *Random sampling*
3) *No perfect collinearity*
4) *Zero-conditional mean*
5) *Homoscedasticity*
6) ***Normality***

Heteroskedasticity

Residual plot u hat v.s. y hat



1) *Linear in parameters*
2) *Random sampling*
3) *No perfect collinearity*
4) *Zero-conditional mean*
5) ***Homoscedasticity***
6) *Normality*

Breusch-Pagan Test

Source	SS	df	MS		
Model	5.64821749	18	.31378986	F(18, 10311)	18.57
Residual	174.198896	10,311	.016894472	Prob > F	0.0000
				R-squared	0.0314
				Adj R-squared	0.0297
Total	179.847114	10,329	.017411861	Root MSE	.12998

Number of obs = 10,330

uhat2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Lng_diff_quad	-.4908019	.0755575	-6.50	0.000	-.6389092	-.3426946
Lat_diff_quad	.024121	.064171	0.38	0.707	-.1016666	.1499087
square_log	-.0315782	.0063983	-4.94	0.000	-.04412	-.0190363
livingRoom	.0092832	.0060742	1.53	0.126	-.0026233	.0211898
livingRoom2	-.0006907	.0010895	-0.63	0.526	-.0028263	.001445
drawingRoom	-.0136014	.0070061	-1.94	0.052	-.0273347	.0001319
drawingRoom2	.0056361	.0024965	2.26	0.024	.0007424	.0105298
bathRoom	-.0338162	.0097128	-3.48	0.001	-.0528552	-.0147771
bathRoom2	.0161978	.002452	6.61	0.000	.0113914	.0210042
constructionTime_diff	-.0013111	.0002063	-6.35	0.000	-.0017155	-.0009066
ladderRatio	-.0013404	.007953	-0.17	0.866	-.0169299	.0142491
elevator	-.0257976	.0041969	-6.15	0.000	-.0340242	-.0175709
fiveYearsProperty	-.0156391	.0027297	-5.73	0.000	-.0209899	-.0102884
subway	.0070068	.0028842	2.43	0.015	.0013533	.0126604
buildingType2	-.1572712	.0654481	-2.40	0.016	-.2855621	-.0289803
buildingType3	-.0051371	.0039253	-1.31	0.191	-.0128315	.0025573
buildingType4	-.0049585	.0042149	-1.18	0.239	-.0132206	.0033036
renovationCondition4	-.0079897	.0026536	-3.01	0.003	-.0131912	-.0027882
_cons	.2971915	.0238423	12.46	0.000	.250456	.343927

Third Iteration

Heteroskedasticity Robust Version - Final Model

Linear regression

Number of obs	=	10,330	
F(18, 10311)	=	471.59	
Prob > F	=	0.0000	
R-squared	=	0.4436	
Root MSE	=	.31608	

price_log	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
Lng_diff_quad	-7.772003	.1794032	-43.32	0.000	-8.123668	-7.420338
Lat_diff_quad	-6.045975	.1498406	-40.35	0.000	-6.339692	-5.752259
square_log	-.3266811	.0171551	-19.04	0.000	-.3603085	-.2930537
livingRoom	.1266209	.0151061	8.38	0.000	.0970101	.1562318
livingRoom2	-.0160333	.0028228	-5.68	0.000	-.0215666	-.0105001
drawingRoom	.1380134	.0178715	7.72	0.000	.1029818	.1730451
drawingRoom2	-.0269155	.0065875	-4.09	0.000	-.0398282	-.0140028
bathRoom2	.0177667	.0033566	5.29	0.000	.0111872	.0243462
constructionTime_diff	.0090475	.0005247	17.24	0.000	.0080189	.0100761
ladderRatio	.1711196	.0219524	7.80	0.000	.1280887	.2141505
elevator	.1469463	.0148498	9.90	0.000	.1178378	.1760549
fiveYearsProperty	-.016668	.006749	-2.47	0.014	-.0298974	-.0034386
subway	.0928612	.0069073	13.44	0.000	.0793215	.1064008
buildingType2	.5371018	.0414938	12.94	0.000	.4557659	.6184377
buildingType3	.0712969	.009512	7.50	0.000	.0526515	.0899423
buildingType4	.0931076	.010597	8.79	0.000	.0723355	.1138797
renovationCondition4	.0832504	.0064659	12.88	0.000	.0705759	.0959248
buildingStructure6	.0563847	.0138637	4.07	0.000	.0292091	.0835602
_cons	11.57876	.0659371	175.60	0.000	11.44952	11.70801

Conclusion

Interpretation of Our Model

price_log=(+11.58)

(-7.77) Lng_diff_quad **(-6.05)**Lat_diff_quad **(-0.33)**square_log

(+0.13)livingroom **(-0.02)**livingroom^2

(+0.14)drawingroom **(-0.03)**drawingroom^2 **(+0.02)**bathroom^2

(+0.01)constructionTime **(+0.17)**ladderRatio **(+0.15)**elevator

(-0.02)fiveYearsProperty **(+0.09)**subway

(+0.54)buildingType2 **(+0.07)**buildingType3 **(+0.09)**buildingType4

(+0.08)renovationCondition4 **(+0.06)**buildingStructure6 **(+u)**

price_log= (+11.58)

(-7.77) Lng_diff_quad (-6.05)Lat_diff_quad (-0.33)square_log

(+0.13)livingroom (-0.02)livingroom^2

(+0.14)drawingroom (-0.03)drawingroom^2 (+0.02)bathroom^2

(+0.01)constructionTime (+0.17)ladderRatio (+0.15)elevator

(-0.02)fiveYearsProperty (+0.09)subway

(+0.54)buildingType2 (+0.07)buildingType3 (+0.09)buildingType4

(+0.08)renovationCondition4 (+0.06)buildingStructure6 (+u)

(-7.77) Lng_diff_quad (-6.05)Lat_diff_quad

price_log= (+11.58)

(-7.77) Lng_diff_quad (-6.05)Lat_diff_quad (-0.33)square_log

(+0.13)livingroom (-0.02)livingroom^2

(+0.14)drawingroom (-0.03)drawingroom^2 (+0.02)bathroom^2

(+0.01)constructionTime (+0.17)ladderRatio (+0.15)elevator

(-0.02)fiveYearsProperty (+0.09)subway

(+0.54)buildingType2 (+0.07)buildingType3 (+0.09)buildingType4

(+0.08)renovationCondition4 (+0.06)buildingStructure6 (+u)

(+0.54)buildingType2 (bungalow/Siheyuan) (base category: tower):

- If the house is a bungalow → price increase by [exp(0.54) - 1] → 71.6% (relative to tower).

price_log= (+11.58)

(-7.77) Lng_diff_quad (-6.05)Lat_diff_quad **(-0.33)square_log**

(+0.13)livingroom (-0.02)livingroom^2

(+0.14)drawingroom (-0.03)drawingroom^2 (+0.02)bathroom^2

(+0.01)constructionTime (+0.17)ladderRatio (+0.15)elevator

(-0.02)fiveYearsProperty (+0.09)subway

(+0.54)buildingType2 (+0.07)buildingType3 (+0.09)buildingType4

(+0.08)renovationCondition4 (+0.06)buildingStructure6 (+u)

(-0.33)square_log:

- For every 1% increase in the total square meter of the house → 0.33% decrease in price.

price_log= (+11.58)

(-7.77) Lng_diff_quad (-6.05)Lat_diff_quad (-0.33)square_log

(+0.13)livingroom (-0.02)livingroom^2

(+0.14)drawingroom (-0.03)drawingroom^2 (+0.02)bathroom^2

(+0.01)constructionTime (+0.17)ladderRatio (+0.15)elevator

(-0.02)fiveYearsProperty (+0.09)subway

(+0.54)buildingType2 (+0.07)buildingType3 (+0.09)buildingType4

(+0.08)renovationCondition4 (+0.06)buildingStructure6 (+u)

(+0.13)livingroom (-0.02)livingroom^2

(+0.14)drawingroom (-0.03)drawingroom^2

(+0.02)bathroom^2

price_log= (+11.58)

(-7.77) Lng_diff_quad (-6.05)Lat_diff_quad (-0.33)square_log

(+0.13)livingroom (-0.02)livingroom^2

(+0.14)drawingroom (-0.03)drawingroom^2 (+0.02)bathroom^2

(+0.01)constructionTime (+0.17)ladderRatio (+0.15)elevator

(-0.02)fiveYearsProperty (+0.09)subway

(+0.54)buildingType2 (+0.07)buildingType3 (+0.09)buildingType4

(+0.08)renovationCondition4 (+0.06)buildingStructure6 (+u)

(+0.01)constructionTime

- To our surprise, the age of construction only has a weak positive effect on price. Unit house price increase by 1% for every 1 year increase in age.

Critical Thoughts

(+0.01)constructionTime

(0) construction Time^2

Hypothesis:

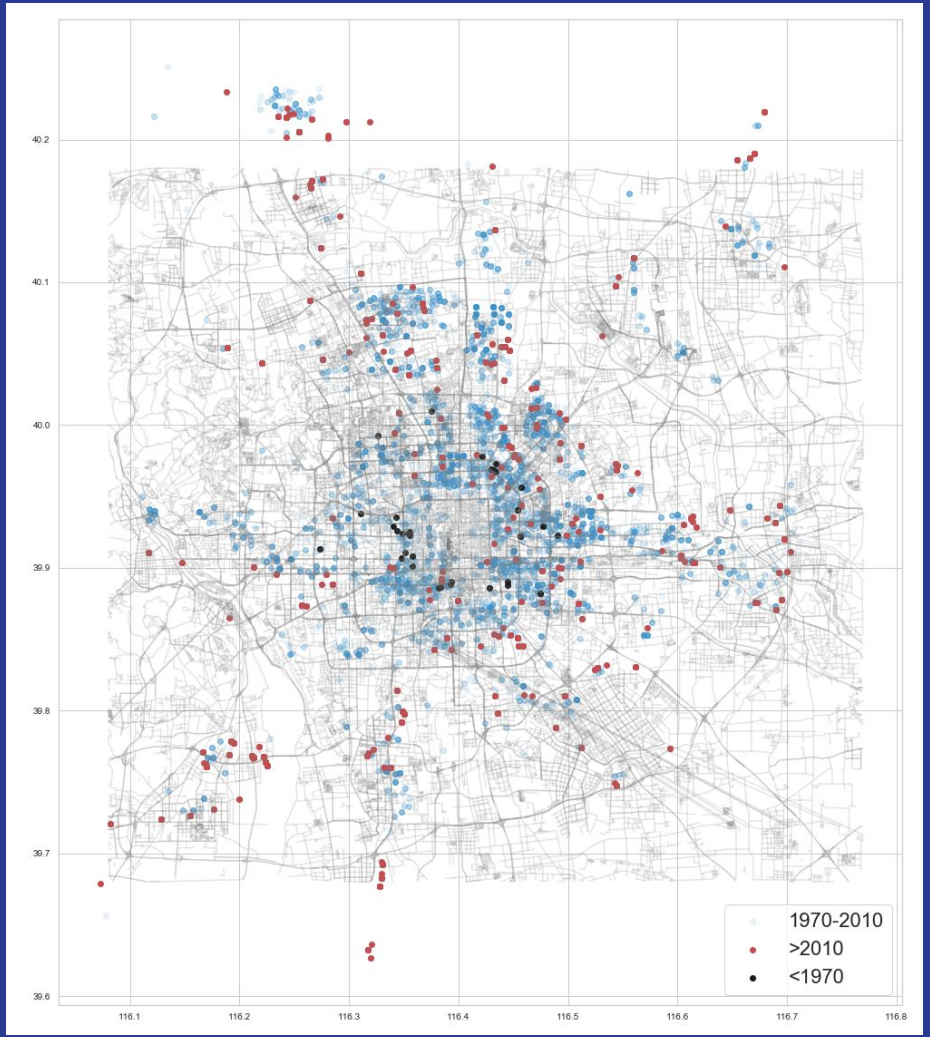
(A) age

+(B) age2

(A<0, B>0)

(+0.01)constructionTime

(0) construction Time^2



Limitations and Further Work

Methodology Limitation and Further Work

Heteroskedasticity

- OLS with robust standard error
- WLS (weighted least squares)
- FGLS (feasible generalized least squares)

Heteroskedasticity

- OLS with robust standard error
- WLS (weighted least squares)
- FGLS (feasible generalized least squares)

Heteroskedasticity

- OLS with robust standard error
- WLS (weighted least squares)
- FGLS (feasible generalized least squares)

Information Limitation and Suggestion

Function forms of the variables

1. Economic theories
    a. Hedonic pricing function (Rosen, 1947)
        i. Structure
        ii. Location
        iii. Neighborhood
2. Common sense

Function forms of the variables

1. **Housing structure:**
    a. Number of floor
    b. Facing of the house
2. **Housing location:**
    a.
    b.
3. **Neighborhood:**
    a.
    b.
    c.
    d.

Function forms of the variables

1. **Housing structure:**
    a. Number of floor
    b. Facing of the house
2. **Housing location:**
    a. Macro: Distance to secondary city-centers (miles)
    b. Micro: Distance to school, hospital, recreational area, etc
3. **Neighborhood:**
    a.
    b.
    c.
    d.

Function forms of the variables

1. **Housing structure:**
    a. Number of floor
    b. Facing of the house
2. **Housing location:**
    a. Macro: Distance to secondary city-centers (miles)
    b. Micro: Distance to school, hospital, recreational area, etc
3. **Neighborhood:**
    a. Median household income
    b. School rating
    c. Number of major crimes
    d. Residential density
    ...

Reference

References

- House price image
- Chen, J. and Hao, Q.J. (2006). "Housing Market Development and Housing Affordability in Shanghai 1993-2005". Paper for Uppsala-Tsinghua Joint Conference on “Housing Affordability in China”, Beijing, April.22-24, 2006.
- Fletcher,M., Gallimore,P. & Mangan,J., (2000). Heteroscedasticity in Hedonic Price Models. Journal of Property Research, 17(2), pp. 93-108.
- John F. Kain & John M. Quigley (1970) Measuring the Value of Housing Quality, Journal of the American Statistical Association, 65:330, pp. 532-548.
- Nicodemo, Catia & Raya, Josep Maria, 2012. "Change in the distribution of house prices across Spanish cities," Regional Science and Urban Economics, Elsevier, vol. 42(4), pp. 739-748.
- Yusof, A. and Syuhaida Ismail. “Multiple regressions in analysing house price variations.” Communications of The IbIMA (2012): 1-9.

Thank you!